The following document summarises current and future Methodology Division (MD) research into key themes and directions. We broadly align these themes with the strategic priorities, capabilities, challenges and opportunities described in the ABS Corporate Plan.

The ABS is undergoing a comprehensive transformation covering the areas of infrastructure, environment, strategy, governance, people and culture. This is in response to drivers that include growing customer expectations, the need for greater efficiency and innovation, and the dynamic information environment in which the ABS now operates. The ABS MD research program is informed by these drivers and the related ABS strategic priorities which are:

DELIVER HIGH QUALITY STATISTICS

- Continue to deliver quality timely statistics on important matters.
- Maximise the value of public data though integration and improved access.

STRENGTHEN OUR PARTNERSHIPS

- Engage and partner to build statistical solutions and capability
- Reduce the burden on respondents and protect the confidentiality of information provided.

DRIVE HIGH PERFORMANCE

- Innovate to facilitate more agile and efficient outcomes.
- Use technology wisely to transform our infrastructure and processes.
- Promote a high performance culture.

With the above in mind we categorise the relevant research projects in MD into the following three themes:

1. Sustainability - reduce costs and respondent burden and improve responsiveness

2. Provide new products and services

3. Maintain quality and manage risk

Underpinning and implicit in each of these themes is the need to innovate and make stronger use of new technologies and methods. While we assign research projects to a single category, projects may reflect aspects of more than one theme.

Research Theme 1 - Sustainability - reduce costs and respondent burden and improve responsiveness

The ABS continues to seek to make greater use of administrative datasets. This includes deriving statistics directly from administrative data, using administrative data as inputs into data integration (see research theme 2), and improving sample design and estimation. A key current example is the adoption of the GNAF-based address register as a population frame for household surveys. This motivates research into dwelling level targeting of subgroups, combining list and area based sampling, conditional selection and opportunities for improved estimation.

Work continues to be progressed in the area of responsive survey design to rationalise survey follow-up while maintaining sample representativeness. Manual formation and allocation of survey enumeration workloads to interviewers has been partially automated by constrained optimisation methods, and further refinements and modifications are desired. MD is also looking at improving contact methods for survey enumeration making stronger use of multiple modes of collection. Related work is around respondent analysis for more effective follow-up actions.

Modelled small domain estimates have been produced in a limited way by the ABS, and there is a desire to extend this in the future. Research currently includes the use of time series models for small domain estimation and deriving better quality measures for simple synthetic estimators. A related small sample estimation problem arises from the plan to transition from periodic to continuous

enumeration for ABS household surveys. Work here is to investigate simple but robust modelling methods for pooling data from repeated surveys.

Research Theme 2 - Provide new products and services

Data integration has been, and will continue to be, a key research area for the ABS. While methods for linkage, analysis and quality measurement of two datasets are becoming mature, there is a need to for further research into multi-way linkage and measuring the associated linkage accuracy. Closely related is the desire to create a linked data 'spine' with the possibility of applying Bayesian entity resolution methods for this purpose. Other research problems in data integration include weighting for longitudinally linked datasets and improved measurement of the overlap between two linked datasets when this is non-trivial.

The ABS is keen to provide greater microdata access to users, while maintaining strong confidentiality protections for unit-level data. The ABS currently allows users a large freedom and flexibility to analyse microdata through datalabs, but with an associated costly exercise in manual checking of the outputs. A more automated approach is desirable that can identify units at risk of spontaneous recognition. MD research is to develop effective distance metrics for 'remarkability.' This includes methods that allow for 'remarkability' in the time dimension, with functional data analysis a potential solution. For tabular releases that may require both primary and secondary suppression of table cells, MD is looking to apply constrained optimisation methods. Work is also being conducted to evaluate the utility of synthetic microdata sets for release to users.

There is potential for the ABS to extract greater value from the large number of unit record and aggregate datasets over which it has custody. ‘Big data’ methods could be used to search for signature patterns in the inter-related datasets in a computationally feasible way. This research has a number of applications, for example, tracking exporters over time (which is problematic using ABN) and evaluating the plausibility of unusual movements in economic time series. Work is also required to understand and account for the statistical properties of streaming data, for example transactional datasets.

Research Theme 3 - Maintain quality and manage risk

With a raft of changes soon to be implemented to systems, processes and methods, there is a need to identify and measure impacts to ABS time series. State space models, including Seemingly Unrelated Time Series models, are being examined for this purpose. These methods need to be applied to both broad and fine domains with the challenge of maintaining strong coherence across series. Improved methods of backcasting are of interest when it is necessary to revise series in time on the basis of the measured impacts. Optimal and practical experimental designs need to be constructed in association with the time series models.

In the ABS output adjustments are sometimes applied to estimates in order to ensure coherence at a point in time (contemporaneous benchmarking) and coherence both contemporaneously and temporally. A key example is that of supply-use tables produced for the National Accounts. Research is being conducted to look at applying constrained optimisation methods to replace the current manual approaches for contemporaneous benchmarking. The multivariate Denton method will be investigated for contemporaneous and temporal benchmarking. Applications of these methods may also apply to demographic outputs and to systems of seasonally adjusted time series.

Many ABS surveys make use of imputation for dealing with missingness. Deterministic regression is sometimes used but this has the potential to cause spurious spikes in the data. Fractional imputation is being considered an alternative. Another quality improvement for imputation involves developing and applying constrained optimisation methods to ensure imputation values are automatically consistent with linear edit rules.